

Data Sharing Between the Bureau of Labor Statistics and the Census Bureau

BLS Data Synchronization Project

Sharon Stang

Economist

Division of Administrative Statistics/
Office of Employment and Unemployment Statistics

FedCASIC 2016

May 4, 2016



A long time ago, in a galaxy far,
far away....

EPISODE I

THE PHANTOM MENACE

Turmoil has engulfed the galactic statistical community. The disparities between NAICS codes on the Census Bureau and Bureau of Labor Statistics business registers is under intense scrutiny.

Hoping to resolve the matter peacefully, a workshop to discuss interagency data sharing was organized by the National Research Council on behalf of the Bureau of Economic Analysis.

History

- CIPSEA 2002
- Data sharing between BEA, BLS, and the Census Bureau (2005)
- Protect confidentiality and improve the quality of data and statistics
- Benefits agencies and data users



EPISODE II

ATTACK OF THE CLONES

While much time has passed, there is lingering unrest in the galactic statistical community. A legislative blockade has made it difficult for the individual agencies to share their administrative data sets for the common good of the Republic.

Seven years after their first meeting, the Census Bureau and the Bureau of Labor Statistics forge a pact to share their multi-unit data, and embark on a quest to rid the galaxy of NAICS coding discrepancies.

2012 - MOU Between BLS and Census

- Share certain business data of multi-location companies.
- Research and fix industry coding discrepancies



Data Sources

■ Census Business Register (BR)

- ▶ 2007/2012 Economic Censuses
- ▶ Annual Company Organization Survey (COS)
- ▶ Annual Survey of Manufacturers (ASM)
- ▶ Service Annual Surveys (SAS)

■ BLS Quarterly Census of Employment and Wages (QCEW)

- ▶ 2007 Annual data and 2011/2012 Q1-Q2 data
- ▶ Quarterly State Unemployment Insurance (UI) filings
- ▶ Annual Refiling Survey (ARS)



Shared Data

■ Census

- ▶ Data free of Title 26 Restrictions
- ▶ For multi-location companies
 - Business name and address
 - State, county, and place geocodes
 - EIN
 - NAICS code
 - NAICS Source Code
 - ❖ 2007 March Employment
 - ❖ 2007 Annual Payroll

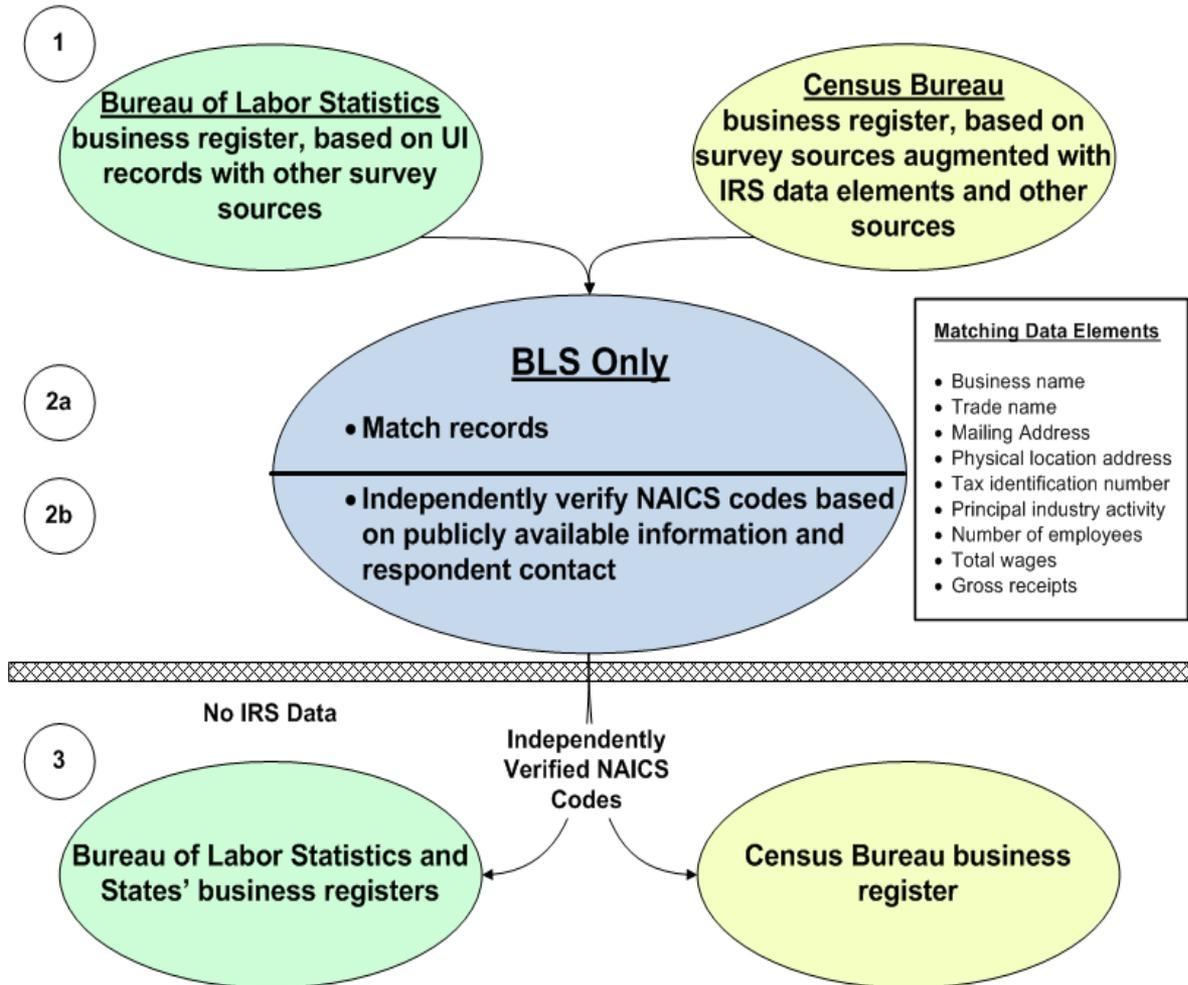
■ BLS

- ▶ Multi-unit company name and address
- ▶ Affiliated establishment information
 - Name and address
 - EIN
 - NAICS code
 - ARS year
 - Operational Status
 - Quarterly Employment
 - Quarterly and annual payroll

❖ 2012 Economic Census data available mid-2014.



BLS Data Synchronization Project



What are Multis?

- Census – Multiple establishments with the same federal Employer Identification Number (EIN).
- BLS – Multiple establishments with the same Unemployment Insurance Account Number (UIN) within a STATE.

QCEW/Census 2012* file matching

Matching Criteria: EIN, State FIPS, Company Name, Physical Location Address (PLA), and County/ Township Code - after standardization.

■ Worksites on the 2012 BR file:	1,915,448
■ Number of matched worksites:	905,853
■ Worksites on the 2012-Q1 QCEW file:	9,247,203
■ Number of matched QCEW worksites:	884,597
■ Total Matches:	857,743
■ One to One matches:	817,209
■ QCEW NAICS = Census NAICS	611,944 (75%)
■ QCEW NAICS <> Census NAICS	205,265 (25%)

❖ 2012 Economic Census data made available in May 2014.



Exclusion Criteria

- Establishments with less than 50 in QCEW employment in March 2012
- Establishments with an absolute employment difference greater than 30% between the March 2012 Census employment and the March 2012 QCEW employment
- Establishments with no sector mismatch between Census and QCEW
 - ▶ **7,790** in-scope establishments
 - 100+ employment, sector mismatch, and Census/QCEW Employment difference $\leq 30\%$ --- **3,820** establishments
 - Employment between 50-99 --- **3,970** establishments (lower priority but in-scope)

Census 2007/2012 File Comparison

- **5,972** originally identified 2007 in-scope establishments
- **7,990** in-scope establishments in 2012
 - ▶ **3,256** (42%) were also in the original 2007 matched establishments.
 - ▶ **4,534** (58%) were newly identified as sector mismatches.

EPISODE III

REVENGE OF THE SITH

Jedi Knights have been dispatched to all corners of the galaxy, to root out the causes of NAICS discrepancies in the data files.

Once identified, the brave Jedi will coordinate a two-pronged attack to resolve these conflicts and return data consistency to the galaxy.

NAICS Adjudication Steps

- BLS regional office staff were provided with the QCEW establishment information only and asked to research, and contact the establishments as necessary, to either verify or update the QCEW NAICS codes.
- A second identical blind review of the establishments was conducted by a different regional office.
- Results of the two independent reviews were then evaluated.

NAICS Adjudication Results

- 1,466 establishments were adjudicated in 2014.
- Results:
 - ▶ **464** matched the QCEW NAICS Sector. Candidates for Census NAICS change.
 - ▶ **247** matched the Census NAICS Sector. Candidates for QCEW NAICS change.
 - ▶ **53** did not match the QCEW or Census NAICS Sectors. Candidates for Census and QCEW NAICS Change.
 - ▶ **702** had different suggested NAICS sectors in both reviews.

What We learned

- Matching microdata records is hard
- Some discrepancies fix themselves over time
- Discrepancies exist because some establishments are just hard to code
- This work is time-consuming and resource intensive
- There will likely always be discrepancies given how the programs are structured



EPISODE IV

A NEW HOPE

Armed with extensive new information on the scope of the NAICS coding discrepancies across the galaxy, the Bureau of Labor Statistics has begun planning its counter-attack.

This macro-level attack heralds an awakening of the Force within the galaxy as Census and the Bureau of Labor Statistics work tirelessly toward the goal of consistent, protected, reconcilable business register data.

BLS Data Synchronization Project: Phase -2

- Take a macro-level approach
- Identify significant sub-sector level differences between the BR and QCEW
- Research why discrepancies exist
- Develop strategies to address discrepancies based on their root causes



NAICS sub-sectors of interest

2012 CBP and QCEW Published Data	Establishments				Employment				Wages (millions)			
	QCEW	CBP	Absolute Value Difference	% Diff	QCEW	CBP	Absolute Value Difference	% Diff	QCEW	CBP	Absolute Value Difference	% Diff
NAICS Sub-Sector	QCEW	CBP	Absolute Value Difference	% Diff	QCEW	CBP	Absolute Value Difference	% Diff	QCEW	CBP	Absolute Value Difference	% Diff
NAICS 561 Administrative and support services	449,229	361,733	87,496	19.48%	7,221,478	9,839,247	2,617,769	36.25%	\$248,773	\$327,316	\$78,543	31.57%
NAICS 611 Educational services	99,704	68,276	31,428	31.52%	2,633,156	653,409	1,979,747	75.19%	\$112,888	\$17,659	\$95,229	84.36%
NAICS 551 Management of companies and enterprises	55,835	52,380	3,455	6.19%	1,983,979	3,065,905	1,081,926	54.53%	\$258,206	\$307,489	\$49,283	19.09%
NAICS 622 Hospitals	8,214	6,546	1,668	20.31%	4,707,397	5,771,797	1,064,400	22.61%	\$260,651	\$317,177	\$56,526	21.69%
NAICS 423 Merchant wholesalers, durable goods	248,211	244,963	3,248	1.31%	2,787,660	3,383,284	595,624	21.37%	\$191,296	\$218,209	\$26,913	14.07%
NAICS 425 Electronic markets and agents and brokers	230,707	44,247	186,460	80.82%	852,594	295,007	557,587	65.40%	\$77,975	\$14,346	\$63,629	81.60%
NAICS 813 Membership associations and organizations	140,214	106,911	33,303	23.75%	1,292,934	885,840	407,094	31.49%	\$48,560	\$38,049	\$10,511	21.64%
NAICS 448 Clothing and clothing accessories stores	129,378	147,502	18,124	14.01%	1,341,516	1,656,842	315,326	23.51%	\$26,690	\$27,864	\$1,174	4.40%
NAICS 517 Telecommunications	43,507	50,576	7,069	16.25%	864,686	1,141,922	277,236	32.06%	\$76,799	\$76,033	\$766	1.00%
NAICS 713 Amusements, gambling, and recreation	71,692	71,562	130	0.18%	1,264,517	1,518,405	253,888	20.08%	\$26,872	\$28,403	\$1,531	5.70%
NAICS 334 Computer and electronic product manufacturing	18,654	13,218	5,436	29.14%	1,095,274	850,613	244,661	22.34%	\$116,448	\$61,802	\$54,646	46.93%
NAICS 518 Data processing, hosting and related services	14,501	14,319	182	1.26%	250,957	486,387	235,430	93.81%	\$23,188	\$40,705	\$17,517	75.55%
NAICS 454 Nonstore retailers	40,085	63,691	23,606	58.89%	431,949	597,414	165,465	38.31%	\$22,900	\$27,012	\$4,112	17.96%
NAICS 511 Publishing industries, except internet	34,599	26,886	7,713	22.29%	735,208	864,723	129,515	17.62%	\$64,661	\$79,884	\$15,223	23.54%
NAICS 213 Support activities for mining	16,386	14,080	2,306	14.07%	383,967	503,950	119,983	31.25%	\$32,736	\$32,236	\$500	1.53%
NAICS 523 Securities, commodity contracts, investments	94,116	98,109	3,993	4.24%	811,024	911,891	100,867	12.44%	\$229,080	\$169,880	\$59,200	25.84%



Contact Information

Sharon Stang

Economist

DASLT/OEUS

www.bls.gov/cew

202-691-5789

stang.sharon@bls.gov

